CLAIMS

What is claimed is:

1. A method of indexing a feature vector space using a tree structure, the method comprising the step of indexing an approximation region in which feature vector elements are sparsely distributed as one special node belonging to a child node of the tree data structure, together with another sparsely distributed approximation region spaced apart by a distance less than a predetermined distance.

2. A method of indexing a feature vector space comprising the steps of:

(a) partitioning the feature vector space into a plurality of approximation regions;

(b) selecting an arbitrary approximation region to determine whether the selected approximation region is heavily or sparsely distributed; and

(c) if the approximation region is determined to be sparsely distributed, indexing the corresponding approximation region as one special node belonging to a child node of the tree data structure, together with any other sparsely distributed approximation region spaced apart by a distance less than a predetermined distance.

3. The method of claim 2, wherein the steps (b) and (c) are repeatedly performed on all approximation regions partitioned in the step (a).

4. The method of claim 2, prior to the step (c), further comprising the step of:

(c-1) if the approximation region selected in the step (b) is determined to be heavily distributed, indexing the corresponding approximation region as an ordinary node, partitioning the corresponding approximation region into a plurality of sub-approximation regions, and repeating the step (b) for the partitioned sub-approximation regions.

5. The method of claim 4, wherein the steps (b) and (c) are performed on all approximation regions partitioned in the step (a).

6. The method of claim 2, after the step (c), further comprising the steps of:

(d) determining whether all approximation regions are indexed as special nodes;

(e) if all approximation regions are not indexed as special nodes, selecting the next approximation region and performing the steps after (b) on the approximation region repeatedly; and

(f) if all approximation regions are indexed as special nodes, completing the indexing.

7. The method of claim 2, wherein the plurality of approximation regions are subspaces used in random indexing.

8. The method of claim 2, wherein the plurality of approximation regions are subspaces used in multi-dimensional scaling (MDS), Fast-map, or locality sensitive hashing

9. The method of claim 2, wherein the step (c) comprises the step of:

(c') if the approximation region is determined to be sparsely distributed, indexing the corresponding approximation region as one special node belonging to a child node of the tree data structure together with an adjacent sparsely distributed approximation region.

10. A method of retrieving a feature vector having features similar to a query vector from a vector space indexed by an indexing method using a tree structure including the step of indexing an approximation region in which feature vector elements are sparsely distributed as one special node belonging to a child node of the tree data structure, together with another sparsely distributed approximation region spaced apart by a distance less than a predetermined distance, the retrieval method comprising the steps of:

(a) determining a special node to which the query vector belongs;

(b) setting the distance between an element of the query vector and an element in an approximation region corresponding to the determined special node, which is the closest to the element of the query vector, as a first threshold value; and

(c) excluding all child nodes of the corresponding node if the distance between the query vector and the approximation region indexed as an ordinary node is greater than or equal to the first threshold value.

11. The method of claim 10, prior to the step (c) further comprising the step of:

(c') selecting an arbitrary node among child nodes of a root node and determining whether the selected node is a special or ordinary node.

12. The method of claim 11, wherein the step (c) comprises the steps of:

(c-1) if the selected node is determined to be an ordinary node in the step (c'), calculating the distance the distance d_{or} between the query vector q and the approximation region nor indexed as the ordinary node according to the following equation:

$$d_{or} = d(q, n_{or}) = \sum_i \begin{cases} |q_i - a_i|^2 \text{ when } q_i \rangle a_i \\ 0 \text{ when } a_i \rangle q_i \rangle b_i \\ |q_i - a_i|^2 \text{ when } b_i \rangle q_i \end{cases}$$

(c-2) determining whether the distance d_{or} between the query vector q and the approximation region n_{or} indexed as the ordinary node is less than the first threshold value;

(c-3) if the distance d_{or} between the query vector q and the approximation region n_{or} indexed as the ordinary node is less than the first threshold value, selecting child nodes of the corresponding node; and

(c-4) if the distance d_{or} between the query vector q and the approximation region n_{or} indexed as the ordinary node is greater than or equal to than the first threshold value, excluding all child nodes of the corresponding node.

13. The method of claim 12, after the step (c-2), further comprising the step of updating the first threshold value with the distance d_{or} if the distance d_{or} is less than the first threshold value.

14. The method of claim 11, after the step (c'), further comprising the step of if the selected node is determined to be a special node in the step (c'), converting a space of approximation region corresponding to the special node into a low-dimensional space.

15. The method of claim 12, after the step (c'), further comprising the steps of:

(c-5) if the node selected in the step (c') is determined to be a special node, converting a space of approximation region corresponding to the node into a low-dimensional space;

(c-6) calculating the distance d_{sp} between the query vector q and each element v in the approximation region n_{sp} indexed as the special node according to the following equation:

$$d_{sp} = d(q, n_{sp}) = \min_{v \in n_{sp}} d(q, v): \text{ and}$$

(c-7) determining elements that satisfy the requirement of the distance d_{sp} being less than the first threshold value to be candidate elements.

16. The method of claim 15, after the step (c-7), further comprising the step of updating the first threshold value with the distance d_{sp} if an element satisfying the requirement of the distance d_{sp} being less than the first threshold value exists.

17. The method of claim 12, after the step (c-4), further comprising the steps of:

determining whether all special nodes have been searched;

selecting the next node to perform the steps after (c-1) repeatedly if all special nodes have not been searched;

determining a predetermined number of elements as finally found elements if all special nodes have been searched.